Flexible risk-based portfolio optimisation *(Github link)*

Emlyn Flint[*], Jayson Landman[†]

2020-07-05

Abstract

The purpose of this study is to present and test a general framework for risk-based investing. It permits various risk-based portfolios such as the global minimum variance, equal risk contribution and equal weight portfolios. The framework also allows for different estimation techniques to be used in finding the portfolios. The design of the study is to collate the existing research on risk-based investing, to analyse some modern methods to reduce estimation risk, to incorporate them in a single coherent framework, and to test the result with South African equity data. The techniques to reduce estimation risk draw from the usual mean-variance and risk-based optimisation literature. The techniques include regime switching, quantile regression, regularisation and subset resampling. In the South African experiment, risk-based portfolios materially outperformed the market weight portfolio out-of-sample using a Sharpe ratio measure. Additionally, the global minimum variance portfolio performed better than other risk-based portfolios. Given the long estimation window, no estimation techniques consistently outperformed the application of sample estimators only.

Keywords: risk-based investing, portfolio optimisaiton, estimation risk.

Contents

1 About

This is a research paper created with the `bookdown` package in R. It is intended to promote reproducibility in academic research.

[*]Legae Peresec
[†]Ninety One Asset Management

2 Introduction

If a risk-averse investor wants to construct portfolios with desirable properties, they would ideally want to find allocations that offer an attractive risk-reward trade-off. Markowitz [1952] developed modern portfolio theory and introduced the mean-variance optimal portfolio as a quantitative solution to this asset allocation problem. However, the reward derived from this portfolio has to be estimated from sample data and is often difficult to accurately predict - which, in turn, leads to Markowitz's mean-variance portfolio being highly sensitive to the estimated portfolio inputs.

As an alternative, risk-based investing provides an avenue for finding portfolios for which expected returns do not need to be estimated, and therefore resolves the portfolio expected return sensitivity problem. Examples of risk-based portfolios commonly seen in practice are the global minimum variance portfolio, the equal risk contribution portfolio, and the equal weight portfolio. These three portfolios are optimal for investors that prioritise weight diversification, risk diversification, or a specific combination of both.

Nevertheless, risk-based portfolios remain sensitive to covariance matrix estimation and hence estimation risk. Improving risk-based estimation is done in three ways in this research. The first improvement alters the covariance estimation procedure by accounting for differences in the sample data. These changes include grouping to account for both non-normality and state-based inhomogeneity. The second involves penalising the optimisation to limit the range of admissable portfolios, which increases the investor's odds of choosing a well-estimated portfolio. The final enhancement changes the implementation methodology entirely by performing the portfolio optimisation on subsets of assets and then resampling to find an aggregate portfolio.

This research aims to bring together useful elements of risk-based portfolio estimation and construction methodology into a single flexible framework. The general structure allows a choice of risk-based portfolio as well as estimation risk reduction technique to improve the out-of-sample portfolio performance. Once we have established a framework, the specific portfolio and estimation technique, examples are developed theoretically. All of these reforms will be hollow without being applied to actual financial data. Therefore, the various estimation techniques and risk-based portfolio pairs are back-tested using South African equity data in an experiment, with the results being measured by standard performance methodologies.

This research is built on the work of several different authors. Firstly - as with nearly all portfolio construction research - this dissertation hinges on the modern portfolio theory of Markowitz [1952]. It then considers the particular case of risk-based investing and makes use of the generalised frameworks introduced by Jurczenko et al. [2013] and Richard and Roncalli [2015]. Finally, in terms of improving the estimation and optimisation processes, we make use of the ideas investigated by Flint and Du Plooy [2018], Kinn [2018], and Shen and Wang [2017].

The rest of this dissertation is set out as follows. Chapter 2 outlines a general framework for constructing risk-based portfolios and estimating them in a robust manner. Chapter 3 gives an overview of the specific risk-based portfolios considered in this work. Chapter 4 presents several techniques for reducing estimation risk, exploring their theoretical underpinnings and providing general intuition. Chapter 5 then considers the empirical application of these techniques, highlighting several technicalities. Chapter 6 then applies the flexible risk-based framework to SA equity data, providing an empirical comparison of different implementations. Chapter 7 concludes the research and provides avenues for further research.

3 A framework for constructing risk-based portfolios

3.1 An overview of modern portfolio theory

Every investor has a universe of N assets to which they can allocate capital. The proportion of their allocation to the i^{th} asset, w_i, depends on the investor's risk and return preferences. They could prefer riskier asset combinations because they require high capital growth, or they could prefer more stable asset combinations that prioritise capital preservation. The column vector $w = [w_1, w_2, \ldots, w_N]^\intercal$ is an $N \times 1$ vector of allocations that define an investor's portfolio. This portfolio is constrained by the investor's capital budget, which may be articulated through the notion of weights. Therefore, all considered portfolios should adhere to the budget constraint: $\sum_i w_i = 1$.

In his seminal paper, Markowitz [1952] attempts to quantify the asset allocation process. Herein he posits that the investor has to make a risk-reward trade-off when considering their portfolio returns, R_{p}, over a pre-determined time horizon $[0, T]$. In his framework, he measures risk with the variance of portfolio returns $\mathbb{V}\text{ar}[R_{\text{p}}] = w^\intercal \Sigma w$, where Σ is the $N \times N$ asset return covariance matrix. Markowitz measures reward by the expectation of portfolio returns $\mathbb{E}[R_{\text{p}}] = w^\intercal \mu$, where μ is the $N \times 1$ vector of expected asset returns. If the investor fixes expected returns to a constant, $\mathbb{E}[R_{\text{p}}] = c$, they encounter the problem of minimising portfolio risk, $\mathbb{V}\text{ar}[R_{\text{p}}]$. The mean-variance optimal (MVO) portfolio w_{mvo} achieves this goal while adhering to the budget constraint and the fixed expected return constraint. Written mathematically, w_{mvo} is the solution to the following Markowitz optimisation:

$$w_{\text{mvo}} = \underset{w}{\operatorname{argmin}} \Big\{ w^\intercal \Sigma w \Big\}, \tag{1}$$

subject to the constraints:

$$\mathcal{C}(w) = \begin{cases} w^\intercal \mu & = c \\ w^\intercal \underline{1} & = 1 \end{cases}.$$

There is a complication when using this framework in practice because Σ and μ are unknown. The investor can only observe a small set of sample returns from the underlying stock processes that use these quantities as inputs. The sample returns, represented by a $N \times T$ matrix $\mathbf{X}$, can be used to estimate Σ and μ. The sample estimates are:

$$\mathbf{S} = \frac{1}{T-1}(\mathbf{X} - \bar{\mathbf{x}}\underline{1}_T^\intercal)(\mathbf{X} - \bar{\mathbf{x}}\underline{1}_T^\intercal)^\intercal, \tag{2}$$

$$\hat{\mu} = \bar{\mathbf{x}}, \tag{3}$$

where $\bar{\mathbf{x}}$ is a $N \times 1$ vector of mean returns. A total of $\frac{N(N+1)}{2}$ distinct parameters are being estimated for the sample covariance matrix (SCM) while N distinct sample expected returns are estimated. $\mathbf{S}$ and $\hat{\mu}$ can be substituted into equation (1) to infer an MVO portfolio, $\hat{w}_{\text{mvo}}$. For each level of portfolio return $c = \hat{w}^\intercal \hat{\mu}$, an estimated portfolio volatility level $\mathbb{SD}[\hat{R}_{\text{p}}] = \sqrt{\hat{w}^\intercal \mathbf{S} \hat{w}}$ is realised. These estimated expected return-volatility pairs induce a frontier in the expected return-volatility plane, which may or may not be close to the frontier implied by the actual inputs Σ and μ. The actual frontier is optimal for the Markowitz framework, and he terms it the 'efficient frontier'.

The above framework and estimation procedure yields a solution to the asset allocation dilemma, but there is still a sensitivity predicament. Michaud [1989] shows that the MVO procedure, as described above, overweights assets with substantial estimated returns $\bar{\mathbf{x}}$. However, these are the same assets that are likely to have been misestimated. Thus, any potential estimation errors are 'maximised' - an undesirable property that makes $\hat{w}_{\text{mvo}}$ a potential liability for the investor to hold.

The MVO procedure is commonly adjusted to reduce sensitivity to the input $\bar{\mathbf{x}}$ in one of two ways. The first is to estimate expected returns in a manner that targets return drivers or factors, which has motivated the rise of factor-based investing [Ang, 2014]. The second adjustment is to remove the dependency on expected return estimates altogether and only construct portfolios based on their risk properties. The latter has inspired the field of risk-based investing and is the focus of this dissertation.

3.2 Introducing risk-based investing

A general risk-based portfolio optimisation programme is given below in equation (4). Removing the expected return constraint and altering the previous MVO optimisation problem for the consideration of a generalised objective risk function $f(\cdot|\mathbf{X})$ yields:

$$w^* = \underset{w}{\operatorname{argmin}} \Big\{ f(w, \Sigma | \mathbf{X}) \Big\}, \tag{4}$$

subject to the constraint:

$$w^\intercal \underline{1} = 1 \ ,$$

where $f(\cdot|\mathbf{X})$ is a risk metric to be minimised. The choice of $f(\cdot|\mathbf{X})$ determines which risk-based portfolio is the optimal solution. Chapter **make chapeter** expounds on the risk-based portfolio types relevant to this research.

The range of feasible portfolios given by equation (4) is practically too general because unlikely single asset weights are still possible. To this end, an investor should apply a weight constraint to limit feasible allocations, ensuring comparability with practical investing. Jagannathan and Ma [2003] show that risk-based long-short portfolios can have extreme weights in practice, which are unlikely to be accepted by an investor.

Additionally, Jagannathan and Ma [2003] also conclude that imposing the long-only investment constraint on US equities leads to improved efficiency for optimal portfolios constructed with the first two sample moments $\hat{\mu}$ and $\mathbf{S}$. Hence, applying the long-only constraint is both statistically appropriate and practically relevant for most investors.

Equally important is that an investor will be reluctant to concentrate their portfolio in a small number of assets. Limiting the maximum single asset allocation to a selected weight α avoids such concentration. These constraints are concurrently expressed as $\{w : 0 \leq w_i \leq \alpha, \ \forall i\}$ and can be added to optimisation (4). In its current form, the developed framework is still somewhat abstract, so it is not obvious how to improve it. Even so, one can always define specific properties that the framework ought to have for there to be a good chance of it operating as intended.

3.3 Improving risk-based portfolios

Hadamard [1923] defines a mathematical problem as ‘well-posed’ if:

1. the solution exists,
2. the solution is unique,
3. the solution is not overly sensitive to small perturbations in inputs.

Well-posed problems are easier to work with and are more stable than ill-posed problems - ones that fall short of the definition. The expected return constraint was previously disregarded for MVO portfolios because the MVO framework often does not meet requirement 3. when the sample mean returns estimate μ. Risk-based portfolios are therefore more ‘well-posed’ than MVO portfolios.

However, there are two common ways in which risk-based portfolios are also ill-posed. The first is if there are very few sample observations; namely, if $T < N$. In this scenario, the covariance matrix is not of full rank and is therefore not invertible, causing non-unique solutions to w^*. The framework is then ill-posed by 2. The second is if small changes in $\mathbf{X}$ result in large deviations of w^*. The framework is then ill-posed by 3. Many researchers such as Jobson and Korkie [1981] and Best and Grauer [1991] have shown that the latter phenomenon is observed in practice, and often persists even if T is much larger than N. To address point (iii) and make the problem well-posed, one needs a measure of sensitivity. We outline a pseudo-derivation of a sensitivity measure below.

Kinn [2018] views the portfolio optimisation problem from a modelling perspective. The returns on the portfolio are modelled directly by an unknown function $g(\cdot)$:

$$R_p = g(\mathbf{X}) + \epsilon, \tag{5}$$

where ϵ has the normal distribution with zero mean and variance ϕ^2. Estimating $g(\cdot)$ is the aim of using the framework, and it is a real-valued function that is not necessarily differentiable or continuous. Each algorithm q refers to a combination of estimation procedure for the inputs, and a computation of a risk-based portfolio using equation (5). Each q specifies an estimate of the function g, denoted $\hat{g}_q$. In the same way that $\hat{w}^*$ (w^* calculated with sample inputs $\mathbf{S}$ and $\hat{\mu}$) can be used to estimate the out-of-sample risk-based portfolio, $\hat{g}_q$ can predict out-of-sample portfolio returns, which are forecasted most accurately by the unobservable function g.

It is necessary to distinguish between the two types of data that are available. The first is historical data comprising of the matrix $\mathbf{X}_0$ and in-sample returns $R_{\text{p},0}$, which combine to form the set $\mathcal{H}_0$. The second is out-of-sample data $\mathbf{X}_1$ and $R_{\text{p},1}$, which combine to form the set $\mathcal{H}_1$. Algorithm q does not utilise the data contained in $\mathcal{H}_1$, which are chronologically realised after the most recent points in $\mathcal{H}_0$. A mean squared error penalty is appropriate to measure the accuracy with which $\hat{g}_q(\mathbf{X}_1)$ (estimated using $\mathcal{H}_0$) predicts the out of sample returns $R_{\text{p},1}$. Kinn [2018] terms the expectation of the out-of-sample mean squared error "generalisation error" (GE), shown mathematically as:

$$GE(\hat{g}_q) = \mathbb{E}[(R_{\text{p},1} - \hat{g}_q(\mathbf{X}_1))^2 | \mathcal{H}_1], \tag{6}$$

where GE is specific to a sample. The actual quantity of interest is the expected performance of q for many potential sample sets, as we are evaluating q's efficacy holistically. This quantity is called the expected generalisation error across all samples, denoted G_q. Furthermore, G_q may be decomposed [1] to reflect a common modelling trade-off between bias and variance:

$$G_q := \mathbb{E}\Big[\mathbb{E}[(R_{\text{p},1} - \hat{g}_q(\mathbf{X}_1))^2 | \mathcal{H}_1] \Big| \mathcal{H}_0\Big], \tag{7}$$

$$= \underbrace{\Big(g(\mathbf{X}_1) - \mathbb{E}[\hat{g}_q(\mathbf{X}_1)|\mathcal{H}_0]\Big)^2}_{\text{squared bias}} + \underbrace{\mathbb{V}\text{ar}[\hat{g}_q(\mathbf{X}_1)|\mathcal{H}_0]}_{\text{variance}} + \underbrace{\phi^2}_{\text{irreducible error}} . \tag{8}$$

The squared bias is the extent to which the expectation of predicted returns differs from the best possible predictor of returns, the correct function $g(\cdot)$. The variance measures the magnitude by which the predicted returns will vary under repeated sampling. By setting $\hat{g}_q(\cdot) = g(\cdot)$ in equation (8) only statistical noise remains; hence, the noise is irreducible. The risk of misestimating $g(\cdot)$ should not include the risk that is retained by even the best estimator. Therefore, estimation risk is considered as the sum of the first two terms only, the squared bias and the variance. An over-fitted algorithm will have high variation for repeated samples. An under-fitted algorithm will have high bias and be consistently poor for repeated samples. The over- and under-fitting trade-off is an example of how the bias-variance trade-off works in practice.

Until now, we have assumed that the estimated portfolio $\hat{w}^*$ from equation (4) is an unbiased estimate of the actual risk metric-minimising portfolio w^* because the choice of $f(\cdot|\mathbf{X})$ determines precisely the type of risk-based portfolio. However, $f(\cdot|\mathbf{X})$ does not precisely determine the estimation risk. Employing a penalty on the objective function introduces bias to reduce estimation risk, *i.e.* hopefully, the squared bias increase does not outweigh the variance decrease. The introduction of the penalty yields an estimated portfolio $\hat{w}^*$ that is consistently closer to w^* than an unbiased portfolio would be. The penalty constraint can be stated as $P(w) \leq s$ and reduces the set of all possible portfolios. If done correctly allocations that are misestimated by the heftiest margins are excluded by this constraint, and allocations that are consistently closer to the actual portfolio remain. The general risk-based framework can now be restated as below to accommodate the penalty using a Lagrangian multiplier approach:

$$w^* = \underset{w}{\operatorname{argmin}} \Big\{ f(w, \Sigma|\mathbf{X}) + \lambda P(w) \Big\}, \tag{9}$$

subject to the constraints:

$$\mathcal{C}(w) = \begin{cases} w^\intercal \underline{1} = 1 \\ 0 \leq w_i \leq \alpha, \ \forall i \end{cases},$$

where $P(\cdot)$ is the penalty function, and λ is the Lagrangian multiplier. Kinn proposes estimating λ in a way that is consistent with the rest of the optimisation. Therefore, we apply the λ estimation that minimises the portfolio specific risk metric using in-sample data. Additionally, by setting $\lambda = 0$, the unpenalised portfolio can still be recovered. Optimisation (9) will be the general risk-based portfolio optimisation going forward.

Improving on the vanilla risk-based optimisation is done in three ways in this research, and these improvements are also common in the literature. Approach one deals with the assumption in equation (5) that ϵ has a

[1] The decomposition is shown by Friedman et al. [2001].

normal distribution, and that the errors through time are independent and identically distributed. If $\hat{g}_q$ fails to approximate g correctly, then the assumption is violated. There could be natural heterogeneity in the data accounted for by g. Flint and Du Plooy [2018] attempt to deal with heterogeneity by changing the estimation procedure of g so that it includes and accounts for potential differences in each observation of sample data. They do this through the application of regimes and quantiles, grouping the input data to increase the accuracy of the estimated model. The second approach adapted from Kinn [2018] has already been shown above and involves penalising the optimisation and setting $\lambda > 0$ in equation (9). The third and final approach requires resampling from the observations in $\mathbf{X}$ for different subsets of assets and hinges on implementation adjustments. Shen and Wang [2017] suggest finding w^* multiple times for different resampled subsets and blending the result afterwards to find an aggregate weight, that hopefully reduces estimation risk. Chapter **insert chapter ref** contains a more detailed exploration of these techniques. Before that, further investigation of desirable risk properties and the portfolios that have them is required to specify $f(\cdot|\mathbf{X})$.

4 Overview of risk-based portfolios

There are many types of risk-based portfolios, for a broader analysis refer to du Plessis and van Rensburg [2017]. In this chapter, we review three risk-based portfolios: the global minimum variance, equal weight and equal risk contribution portfolios. The risk properties that these portfolios have, the strategies that bear them out and the conditions under which they perform optimally are all covered.

4.1 Risk-based portfolio types

Each risk-based portfolio is optimal in some sense because they all minimise a specific objective risk function. The nature of the objective function depends on the investor's risk preferences. In line with Markowitz [1952], a genetic risk metric to minimise is the portfolio variance, as an investor may not be willing to tolerate large swings in capital value. Setting $f(\cdot|\mathbf{X}) = w^\intercal \Sigma w$ in portfolio optimisation (9) yields the global minimum variance (GMV) portfolio, denoted w_{gmv}. The objective function is the same as for the MVO portfolio for an imputed value of c, meaning that the GMV portfolio sits on the efficient frontier. The GMV portfolio is also the only risk-based portfolio that is always on the efficient frontier.

Maillard et al. [2010] utilise the concept of an asset's marginal risk contribution (MRC) in a portfolio to perform risk-based portfolio calculations. It is the sensitivity of the portfolio volatility to the weight of an asset in a portfolio. Alternative representations of the MRC and an outline of why the weighted sum of the MRC's is minimised for the GMV portfolio are shown in appendix **insert appendix ref**. Below is the mathematical definition of the MRC for the i^{th} asset:

$$\frac{\partial}{\partial w_i}\Big[\sigma(w)\Big] = \frac{(\Sigma w)_i}{\sqrt{w^\intercal \Sigma w}}. \tag{10}$$

When taking the realised portfolio returns as the single risk factor, there is no idiosyncratic risk present for assets in the portfolio against this factor. Accordingly, the GMV portfolio is the lowest possible beta portfolio. The GMV portfolio is, therefore, optimal for the investor that always takes on the least risk at the margin. However, risk has to be estimated, so the investor only knows what the least risky options available to them are in a historical sense. The downside is that always choosing the lowest marginal risk contributing asset is implicitly displaying a high level of confidence in estimations of risk.

Contrastingly, the investor may be at a loss when estimating risk. The only measure of diversification for such an investor is weight diversification. To minimise their risk taken at the margin, this investor would hold the smallest possible weight in each of the assets while still satisfying the constraints. This strategy ensures that the investor avoids the maximum marginal risk contributing asset to the greatest extent possible given that they are unsure which of the N assets it is. The least weight concentrated portfolio that this strategy alludes to is the equal weight (EW) portfolio:

$$w_{ew} = \Big[\frac{1}{N} \ \cdots \ \frac{1}{N}\Big]^\intercal, \tag{11}$$

where the weight diversification measure that this portfolio maximises is the inverse Herfindahl index (IHI), calculated as $H^{-1}(w) = (\sum_{i=1}^{N} w_i^2)^{-1}$. The most weight concentrated portfolio is the single asset portfolio - where a single non-zero asset weight is 1. The IHI of this portfolio is 1, which is as low as it can be. On the other hand, the EW portfolio has an IHI value of N. All portfolios, therefore, have an IHI on the interval $[1, N]$. In the presence of a maximum weight constraint, the lower bound of the interval changes. The new lower bound is derived in appendix **insert appendix ref**. To find the EW portfolio using framework (9), set the objective function to the Herfindahl index, $f(\cdot|\mathbf{X}) = \sum_{i=1}^{N} w_i^2$. No parameters in the objective function require estimation, so this is a sample return independent optimisation, which reflects the investor's lack of confidence in historical data.

The GMV and EW portfolios represent two extremes of investors, specifically those that value volatility reduction only and those that value weight diversification only. The inequality $\sigma(w_{gmv}) \leq \sigma(w^{\diamond}) \leq \sigma(w_{ew})$ verifies this intuition. The portfolio $w^{\diamond}$ has intermediate weight concentration. Proof of the inequality is given in appendix **insert appendix ref**. It is unlikely that a capital allocator will completely disregard one risk-based approach for another. One method to construct a sound intermediate portfolio that incorporates the MRC philosophy from the GMV portfolio construction and the high risk-asset avoidance philosophy from the EW portfolio construction is to equalise the total risk contribution (TRC) from each asset. An asset's TRC is the product of its MRC and its weight in the portfolio:

$$\begin{aligned}\text{TRC}_i &= w_i \cdot \frac{\partial}{\partial w_i}\Big[\sigma(w)\Big],\\ &= \frac{w_i \cdot (\Sigma w)_i}{\sqrt{w^{\intercal}\Sigma w}},\end{aligned}$$

where TRC's sum to the portfolio volatility. An equal risk contribution (ERC) portfolio equalises all of the TRC's so that no single asset is a comparatively significant contributor to risk. The choice of $f(\cdot|\mathbf{X})$ that minimises the squared distances between the TRC's to the greatest extent possible is given below:

$$f(w, \Sigma|\mathbf{X}) = \sum_{i=1}^{N}\sum_{j\geq i}^{N}(w_i(\Sigma w)_i - w_j(\Sigma w)_j)^2 \ . \tag{12}$$

Maillard et al. [2010] show that a log-constraint on the weights in GMV optimisation could equivalently express this choice of $f(\cdot|\mathbf{X})$ - an idea explored further in appendix **insert appendix ref**. Therein it is shown that the ERC portfolio is an intermediate portfolio $w^{\diamond}$. While there are several other risk-based portfolios that have been suggested in the literature, we will focus our attention only on these three portfolios, which are arguably some of the most common risk-based portfolios seen in practice Jurczenko et al. [2013].

4.2 Risk-based portfolio properties

Section 3.1 introduces the Markowitz efficient frontier.@S64 extends this work to deduce that there is an optimal portfolio called the tangency portfolio. He does make certain assumptions about investors' preferences and the presence of a risk-free asset. The market-weighted (MW) portfolio is the portfolio for which the efficient frontier is tangential to the line bisecting the y-axis at the risk-free rate (r_f) in the expected return-volatility plane. The MW portfolio is the portfolio held by all investors in the market on average and is relevant because it offers the investor diversification with negligible transaction costs Perold [2007]. The MW portfolio is not risk-based in the traditional sense, but it does not require an estimate of expected returns to calculate; hence, the MW portfolio offers a cheap benchmark against which to compare risk-based portfolio performance. However, the holder of the MW does implicitly adopt all investors' weighted expectations of expected returns Haugen and Baker [1991]. The tangency portfolio is optimal for the Sharpe ratio (SR) measure under Sharpe's assumptions. The measure is defined as:

$$\text{SR}_\text{p} = \frac{\mathbb{E}[R_\text{p}] - r_\text{f}}{\mathbb{SD}[R_\text{p}]} \ . \tag{13}$$

Within the MVO construction, the MW portfolio has the maximum Sharpe ratio (MSR) and is, therefore, the MSR portfolio. Scherer [2007] shows that the MSR portfolio, w_{msr}, can alternatively be expressed as the portfolio for which marginal excess returns and the MRC's are equal for all portfolio constituents. Jurczenko et al. [2013] use this fact to find MSR optimality conditions for each of the risk-based portfolios, some examples of which are shown in appendix **insert appendix num**.

Table 1 summarises the salient risk properties of the EW, ERC, and GMV portfolios. Included is the strategy to find the portfolio, the requirements for when the portfolio coincides with the MSR portfolio, and their empirical risk characteristics. The risk characteristics entail whether the risk is inherent to the investment, the construction of the portfolio, or liquidity restrictions when creating the portfolio.

Table 1: Risk-based investing portfolio properties Jurczenko et al. [2013]

Portfolio	Strategy	MSR conditions	Risk characteristics
EW	Equalise w_i	Identical excess returns. Identical volatilities. Identical correlations.	Medium to high risk. Insensitive to Σ. Low turnover.
ERC	Equalise TRC_i	Identical Sharpe ratios. Identical correlations.	Medium risk. Moderately sensitive to Σ. Medium turnover.
GMV	Equalise MRC_i	Identical excess returns.	Lowest risk. Highly sensitive to Σ. High turnover.

While the MSR conditions are theoretically compelling, out-of-sample optimality is harder to determine in practice. Haugen and Baker [1991] show that portfolios that are superior to the MW portfolio exist when: short-selling is restricted, investments are taxed, and foreign investors are active market participants. These portfolios should have the same expected return as the MW portfolio with lower volatility. Their statement is true even in an 'efficient market'. Studies of the historical performance show that some portfolios outperform others. In these studies, the authors restrict the asset universe to US equities; hence, their results will not necessarily translate to South Africa. The hope of introducing risk-based portfolios is to find Haugen and Baker's superior portfolios. Evidence supporting this ambition exists. DeMiguel et al. [2007] demonstrate the robust out-of-sample performance of EW portfolios when compared to MVO and MW portfolios for a broad range of asset universes. Clarke et al. [2006] also demonstrate that GMV portfolios show outperformance against the MW and MVO benchmarks. They initially attribute this to the diachronic persistence of covariances when compared to expected returns. In a later paper, Clarke et al. [2011] suggest that the outperformance is due to a bias inherent in the portfolio construction towards stocks that do not move with the rest of the market, but that still have comparatively high expected returns.

Within risk-based portfolios, Kritzman et al. [2010] have shown that GMV portfolios outperform EW portfolios when the implementer uses a long enough estimation window. Therefore, they establish a defence for using optimisation on a sample covariance matrix. This research remains consistent with these findings, using the EW portfolio as a benchmark in pursuit of better out-of-sample performance within the GMV and ERC frameworks. In the next chapter, we outline the techniques used to achieve this aim.

5 Estimation risk reduction techniques

As stated in chapter 3, estimation risk is comprised of squared bias and variance. There are many methods to approach reducing estimation risk, but in this research, we introduce three ways that are consistent with the general risk-based investing framework presented in equation (9). The first method deals with improving the estimation of the inputs to the risk-based portfolio function $f(w, \Sigma|\mathbf{X})$, accounting for heterogeneity in the input data. The second method involves penalising the optimisation objective function to obtain a portfolio estimate with consistently lower deviation from the actual out-of-sample risk-based portfolio solution. The final process entails changing the implementation method in a manner that reduces estimation risk. Every risk reduction technique falls into one of these three categories.

5.1 Improving optimisation inputs

The first approach to improving on the sample ERC and GMV portfolios involves finding better estimates for the input Σ, given the set of sample returns. As stated in equation (5), the function g deals with the sample returns in a manner that ensures the irreducible error, ϕ^2, is independent and identically normally distributed. However, because g is unobservable, our estimation $\hat{g}$ might not ensure this property. Heterogeneity of sample errors for investment portfolios has been observed in empirical finance by Ang and Chen [2002], who demonstrate that negative stock price correlations are less pronounced in downward markets. Therefore, empirical finance suggests two states of the world, one where the market is in turmoil, and one where the market is not.

Kritzman et al. [2012] use these two states to determine separate multi-asset allocations for 'turbulent' and 'quiet' markets and adopt a regime switching (RS) approach. To define two regimes, a metric to measure turbulence is required. The authors take a squared Mahalanobis distance (SMD) approach to determine an index through time. The Mahalanobis distance is a multi-dimensional generalisation of the notion of how many standard deviations a point is away from the mean of a distribution. The SMD index (d_t) is expressed mathematically as:

$$d_t = (R_t - \mu)\Sigma^{-1}(R_t - \mu)^\intercal, \tag{14}$$

where:

$$d_t = \mu_{s_t} + \sigma_{s_t}\epsilon_t,$$

and ϵ_t has a standard normal distribution. The state at time t is shown by the random variable s_t. As asserted earlier, the two states are Quiet (Q) and Turbulent (T), hence $s_t \in \{Q, T\}$. To calculate the SMD, we require the unobservable inputs Σ and μ. They can be replaced by their sample counterparts, $\mathbf{S}$ and $\hat{\mu}$, to yield $\hat{d}_t$. The SMD has a state-specific mean and volatility; hence different values are observed based on the current system state. If the system emits large values of $\hat{d}_t$, the probability of being in a turbulent market is high. If the system emits small values of $\hat{d}_t$, the likelihood of being in a quiet market is high. The ζ^{th} quantile of the sample SMDs is the point where the system of reference changes. The market state is also an unobservable variable, so the above model is referred to as a hidden Markov model (HMM). The HMM used in this investigation is depicted in figure 1.

The transition matrix stores the probabilities of transition from a state at time t to another at time $t+1$ for ease of computation. It is mathematically shown as:

$$P_{t,t+1} = \begin{bmatrix} p_{QQ} & p_{QT} \\ p_{TQ} & p_{TT} \end{bmatrix},$$

where the matrix applies to all times t. Because the matrix applies to all times the system is called stationary, and the long-run probabilities of being in each state will converge. Once we have determined the most likely state at time t using an algorithm such as the Viterbi algorithm, we can use the series of estimated states $\{\hat{s}_t : t \in \{0, 1, ..., T\}\}$ to partition the data history. The two datasets would be data used for the quiet sample covariance matrix and data used for the turbulent sample covariance matrix. Flint and Du Plooy [2018] blend these two sample matrices using the investor's risk preferences and the most recent probabilities of being in each state, yielding a more sophisticated estimator of Σ.

An alternative approach to dealing with heterogeneity is to focus on the assumption that errors for return forecasting models have a normal distribution. But first, we need to define a general return forecasting model. If the return of a portfolio is viewed through a set of return drivers or risk factors, then returns could be explained in part by those factors and the portfolio's sensitivity to them. Meucci [2010] encapsulates this idea in his asset return model given below:

$$R = \alpha + \beta^\intercal \mathcal{F} + \epsilon, \tag{15}$$

where R is a vector of asset returns (not portfolio returns through time as shown previously), α is a forecastable vector of returns unique to each security, β is a matrix of sensitivities to risk factors, $\mathcal{F}$ is a vector of factors, and ϵ is the error vector. The error vector is assumed to have a normal distribution. Ang and Chen [2002]

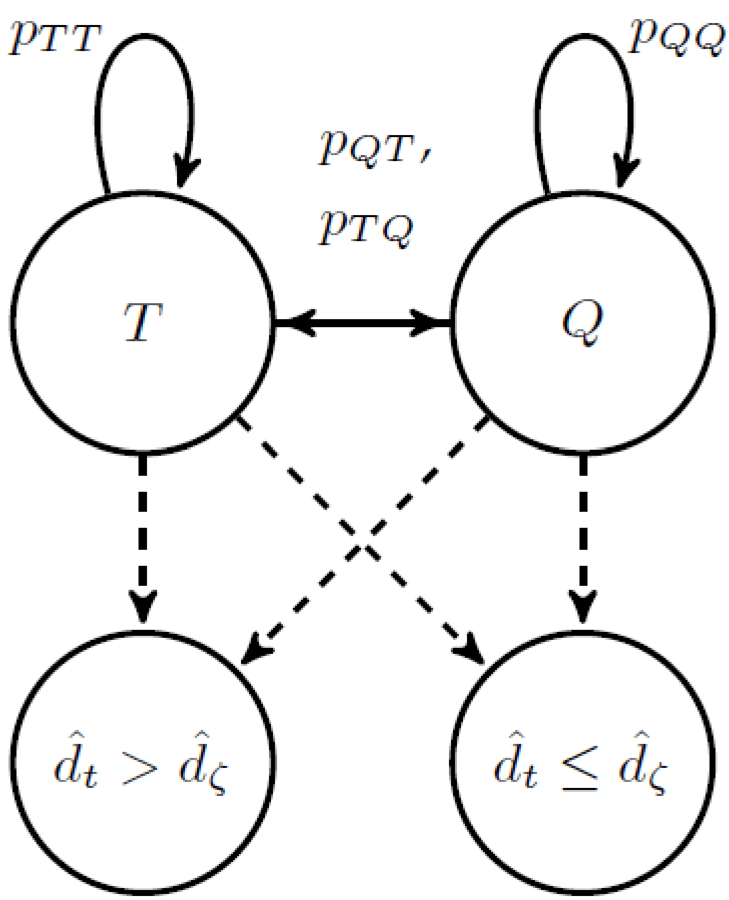

Figure 1: Turbulent / quiet hidden Markov model.

show that in a downward market, the correlation structure is significantly different from what is implied by a normal distribution, which is a problem when using model (15) in the exhibited way. Chen et al. [2019] address the issue of non-normal errors by utilising the quantile regression model proposed by Koenker and Bassett Jr [1978]. The asset returns, idiosyncratic asset returns, asset factor sensitivities and errors could all be considered to be a function of the current quantile, denoted τ. This leads to the quantile factor model (QFM):

$$\mathcal{Q}(\tau) = \alpha(\tau) + \beta(\tau)\mathcal{F} + \epsilon(\tau), \tag{16}$$

where $\tau \in [0, 1]$. In a different symmetric, normally distributed world, τ can be set to 0.5 and model (15) will be recovered. However, in the real world where symmetry and normality are often not adhered to, the quantile conditional errors can be defined more generally so that they only have to satisfy:

$$\mathbb{P}\Big[\epsilon(\tau) \leq \underline{0} \Big| \mathcal{F}\Big] = \tau \ . \tag{17}$$

This structure emerges from the cumulative distribution function (CDF) conditional on the set of factors of each asset return R_i. Given the conditional CDF for the returns on asset i, $F_i(R_i|\mathcal{F})$, the quantile specific inverse CDF, $F_i^{-1}(\tau|\mathcal{F})$, can be used to generate the quantiles $\mathcal{Q}_i(\tau)$. Flint and Du Plooy [2018] suggest using the information about each quantile to construct a series of quantile-specific covariance matrices, which can then be blended to yield a more sophisticated estimator of Σ. Chapter **insert ref to next chapter** covers the implementations of these two techniques to better estimate Σ.

5.2 Penalising the optimization

To add a penalty term in a way that preserves the goal of the risk optimisation, we first need to adapt the objective functions of each risk-based portfolio as given earlier in chapter @ref{rbportch}. Consider the return-targeting penalised optimisation approach of Kinn [2018], both choices of $f(\cdot|\mathbf{X})$ for the GMV and ERC portfolios can be adapted into this approach. Beginning with the GMV portfolio, Kinn views the portfolio variance as an expectation:

$$\begin{aligned} f(w, \Sigma|\mathbf{X}) &= w^\intercal \Sigma w \\ &= w^\intercal(\mathbb{E}[r_t r_t^\intercal] - \mu\mu^\intercal)w && \text{(alternate definition of } \Sigma) \\ &= \mathbb{E}[|w^\intercal\mu - w^\intercal r_t|^2], \end{aligned}$$

where r_t represents the asset returns above the risk-free rate, and μ is a vector of the population expected excess returns as before. Rewriting the portfolio expected excess return as $\bar{r} = w^\intercal\mu$, the idea of return-targeting for a portfolio can be incorporated as the expectation of $|\bar{r} - w^\intercal r_t|^2$, which is the squared distance to a target return level. Kinn's approach is consistent with an MVO optimisation intuitively because the target return level is analogous an expected return constraint and minimising the return's squared distance to this constraint is analogous to variance minimisation. The objective function can now be approximated using the sample average as a result of the law of large numbers. We still have to show how to find the GMV portfolio from an MVO procedure. As stated in table 1, the GMV portfolio is the MSR portfolio if the assumption of identical excess returns is met. Therefore, if the target return is set to a value that is easily obtainable $\bar{r} = \bar{r}_{\text{gmv}}$, then the scheme will yield a GMV portfolio. This easily obtainable value has to be found numerically and cannot be determined a priori. The non-rigorous argument turns out to be empirically true for the portfolios analysed in this research. When the return vector is replaced by the set of sample returns $\mathbf{X}$, and the expectation is approximated by the sample average, the Kinn [2018] form of objective function is recovered:

$$f_{\text{Kinn}}(w|\mathbf{X}) = \frac{1}{T}\sum_{t=1}^{T}(\bar{r}_{\text{gmv}} - \mathbf{X}_t^\intercal w)^2, \tag{18}$$

where t indexes columns of the sample returns matrix $\mathbf{X}$. The GMV portfolio can be found equivalently in this way.

The log-constraint[2], $\sum_i \ln(w_i) \geq c$, can be placed on optimisation (18) to recover the ERC portfolio. To include the constraint in framework (9), a Lagrangian multiplier approach can be used to move the log-constraint to the objective function. Kinn's adapted ERC objective function is then:

$$f_{\text{Kinn}}(w|\mathbf{X}) = \frac{1}{T}\sum_{t=1}^{T}(\bar{r}_{gmv} - \mathbf{X}_t^{\intercal} w)^2 - \eta_{erc}\sum_{i=1}^{N}\ln(w_i), \tag{19}$$

where η_{erc} is the Lagrangian multiplier scalar. Now that we have shown the standard objective functions from earlier are equivalent to the Kinn framework, we have also vindicated the general framework (9) as accommodative of a valid application of supervised machine learning (SML) to portfolio optimisation.

Because logical choices for $f_{\text{Kinn}}(\cdot|\mathbf{X})$ have been established, different penalty functions can be applied to the optimisation. Two common penalised regression techniques are lasso regression and ridge regression (RR). In the presence of a long-only constraint, as is applied in this research, a lasso regression does not make sense, because the penalty function is simply the sum of the absolute weights: $P(w) = \sum_{i=1}^{N} |w_i| = 1$. This penalty is equal to 1 for all constrained portfolios. Separately, the RR is obtained by specifying the penalty as the sum of squares for the portfolio weights: $P(w) = \sum_{i=1}^{N} w_i^2$. The penalty reduces the number of admissable concentrated portfolios and intuitively is not unlike incorporating some of the EW portfolio into the ERC or GMV portfolios. Ignoring constraints, Ledoit and Wolf [2004] show that RR has the same effect as shrinking the sample covariance matrix towards the identity matrix for the GMV optimisation:

$$\mathbf{S}_{RR} = \mathbf{S} + \frac{\lambda}{T}\mathbf{I}, \tag{20}$$

where λ is the shrinkage intensity, and T is the number of sample observations. In the presence of constraints, the actual scaling factor is slightly different from Ledoit and Wolf's calculations, but the intuition of shrinkage towards the identity matrix still applies. If λ becomes very large, the minimum variance portfolio will tend towards the EW portfolio. Estimating lambda is thus a practical choice, and the process to do so consistently is outlined in the next chapter.

5.3 Alternate implementation methods

Shen and Wang [2017] present a means to find a resampled MVO portfolio that reduces estimation risk by optimising random subsets of assets in the investment universe. The process is called subset resampling (SRS). They then aggregate resultant optimised subset portfolios to create a final 'optimal' solution. The procedure requires the inputs of a sample return matrix $\mathbf{X}$ and an asset subset size b. The subset size is related to the extent of the trade-off between bias and variance. We have to choose the degree of repeated sampling, s, which is restricted by the available computational power.

This method can be described as follows. For each of the s repeated samples, we randomly select the j^{th} subset of b assets from the N assets in the investment universe, denoted $\mathcal{I}_j$. Using only the sample return data from the selected asset subset $\mathbf{X}_j$, we then compute the associated optimal portfolios $\hat{w}_j$ using framework (9) and a given choice of objective function. Finally, we average the s optimal subset weight vectors to obtain the final optimal asset portfolio $\hat{w}_{\text{srs}} = (\sum_{j=1}^{s} \hat{w}_j)s^{-1}$.

The SRS process is very general, and could even be applied in conjunction with a penalised optimisation or an improved sample covariance matrix. Additionally, the user can choose the input b. If $b = N$, then the usual sample risk-based portfolio is recovered, albeit in a computationally expensive manner. If $b = 1$, then the SRS procedure will yield the EW portfolio for a large enough value of s. Therefore, b is the input parameter controlling the extent of the trade-off between weight diversification and estimation risk. The estimation of b should be done in a manner consistent with the aim of the optimisation. To ensure b scales with the size of the asset universe, Shen and Wang [2017] recommend writing it in the form $b = N^{\alpha}$, where $\alpha \in [0, 1]$.

The SRS method is comparable to ensemble methods in machine learning. The logical basis is that many different models can be used and aggregated into a final model, rather than assuming a single model is the

[2] The log-constraint is also introduced in appendix **insert appendix ref**.

most accurate to use. Despite the general nature of the SRS procedure, it is still consistent with the approach of increasing the squared bias out of the hope that the variance reduction will offset it enough to lessen overall estimation risk.

Each of the three estimation risk reduction classes has at least one specific technique within them. Some modelling decisions need to be made for a prospective user to apply these techniques in an experiment. These modelling decisions are covered in the next chapter.

References

Andrew Ang. *Asset management: A systematic approach to factor investing.* Oxford University Press, 2014.

Andrew Ang and Joseph Chen. Asymmetric correlations of equity portfolios. *Journal of financial Economics*, 63(3):443–494, 2002.

Michael J Best and Robert R Grauer. Sensitivity analysis for mean-variance portfolio problems. *Management Science*, 37(8):980–989, 1991.

Liang Chen, J. Juan Dolado, and Jesus Gonzalo. Quantile factor models. Working paper, 2019.

Roger Clarke, Harindra De Silva, and Steven Thorley. Minimum-variance portfolio composition. *The Journal of Portfolio Management*, 37(2):31–45, 2011.

Roger G Clarke, Harindra De Silva, and Steven Thorley. Minimum-variance portfolios in the us equity market. *The Journal of Portfolio Management*, 33(1):10–24, 2006.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953, 2007.

Hannes du Plessis and Paul van Rensburg. Diversification and the realised volatility of equity portfolios. *Investment Analysts Journal*, 46(3):213–234, 2017.

Emlyn James Flint and Simon Du Plooy. Extending risk budgeting for market regimes and quantile factor models. *Available at SSRN 3141739*, 2018.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

J Hadamard. *Lectures on Chacy's Problem in Linear Partial Differential Equations.* Yale University Press, 1923.

Robert A Haugen and Nardin L Baker. The efficient market inefficiency of capitalization–weighted stock portfolios. *The Journal of Portfolio Management*, 17(3):35–40, 1991.

Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683, 2003.

J David Jobson and Robert M Korkie. Putting markowitz theory to work. *The Journal of Portfolio Management*, 7(4):70–74, 1981.

Emmanuel Jurczenko, Thierry Michel, and Jerome Teiletche. Generalized risk-based investing. *Available at SSRN 2205979*, 2013.

Daniel Kinn. Reducing estimation risk in mean-variance portfolios with machine learning. *arXiv preprint arXiv:1804.01764*, 2018.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Mark Kritzman, Sébastien Page, and David Turkington. In defense of optimization: the fallacy of 1/n. *Financial Analysts Journal*, 66(2):31–39, 2010.

Mark Kritzman, Sebastien Page, and David Turkington. Regime shifts: Implications for dynamic strategies (corrected). *Financial Analysts Journal*, 68(3):22–39, 2012.

Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

Sébastien Maillard, Thierry Roncalli, and Jérôme Teïletche. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70, 2010.

Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

Attilio Meucci. Factors on demand: building a platform for portfolio managers, risk managers and traders. *Risk*, 23(7):84–89, 2010.

Richard O Michaud. The markowitz optimization enigma: Is 'optimized'optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

André F Perold. Fundamentally flawed indexing. *Financial Analysts Journal*, 63(6):31–37, 2007.

Jean-Charles Richard and Thierry Roncalli. Smart beta: Managing diversification of minimum variance portfolios. In *Risk-Based and Factor Investing*, pages 31–63. Elsevier, 2015.

Bernd Scherer. Can robust portfolio optimisation help to build better portfolios? *Journal of Asset Management*, 7(6):374–387, 2007.

Weiwei Shen and Jun Wang. Portfolio selection via subset resampling. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.